08 - Comparison of Several Means

Junvie Pailden

SIUE, F2017, Stat 589

September 14, 2017

Paired Comparisons

- Measurements are often recorded under different sets of experimental conditions to see whether the responses differ significantly over these sets.
- e.g. The efficacy of a new drug or of a saturation advertising campaign.
- Compare two treatments, or the presence and absence of a single treatment, and assign both treatments to the same or identical units/individuals.
- The paired responses may be analyzed by computing their differences, thereby eliminating much of the influence of extraneous unit-to-unit variation.

Wastewater Treatment Example

- Municipal wastewater treatment plants are required by law to monitor their discharges into rivers and streams on a regular basis.
- Concern about the reliability of data from one of these self-monitoring programs led to a study in which samples of effluent were divided and sent to two laboratories for testing.
- One-half of each sample was sent to the Wisconsin State Laboratory of Hygiene, and one-half was sent to a private commercial laboratory routinely used in the monitoring program.
- Measurements of biochemical oxygen demand (BOD) and suspended solids (SS) were obtained, for $n = 11$ sample splits, from two laboratories.
- Do the two laboratories' chemical analyses agree? If differences exists, what is their nature?

```r
waste <- read.table("T6-1.dat")
colnames(waste) <- c("BOD1","SS1","BOD2","SS2")
str(waste)
```

```
# 'data.frame': 11 obs. of  4 variables:
#  $ BOD1: int  6 6 18 8 11 34 28 71 43 33 ...
#  $ SS1 : int  27 23 64 44 30 75 26 124 54 30 ...
#  $ BOD2: int  25 28 36 35 15 44 42 54 34 29 ...
#  $ SS2 : int  15 13 22 29 31 64 30 64 56 20 ...
```

```r
BOD <- data.frame(Lab = c(rep(1,11), rep(2,11)),
                  value = with(waste, c(BOD1, BOD2)))
SS <- data.frame(Lab = c(rep(1,11), rep(2,11)),
                 value = with(waste, c(SS1, SS2)))
```

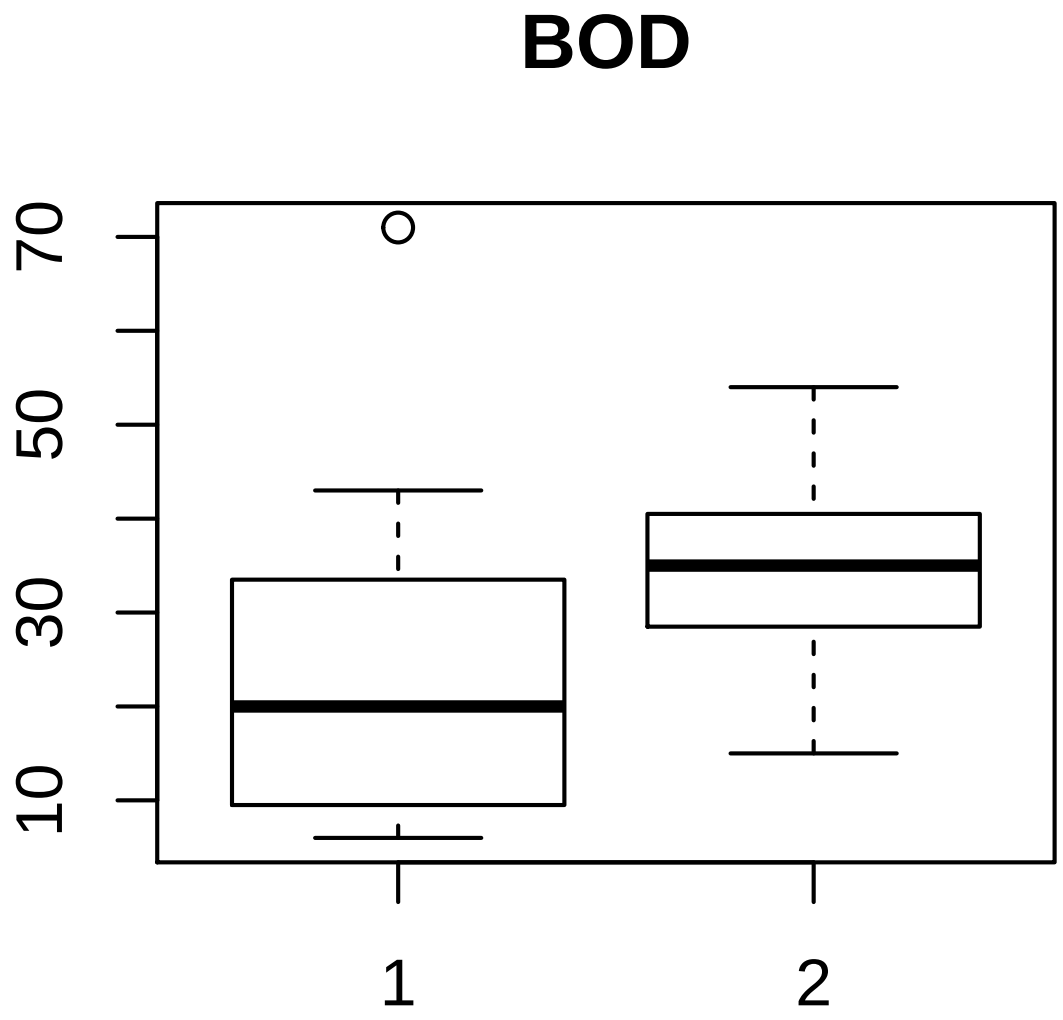
```r
head(BOD)
```

```
#   Lab value
# 1   1     6
# 2   1     6
# 3   1    18
# 4   1     8
# 5   1    11
# 6   1    34
```

```r
tail(BOD)
```

```
#    Lab value
# 17   2    44
# 18   2    42
# 19   2    54
# 20   2    34
# 21   2    29
# 22   2    39
```

```
boxplot(value ~ Lab, data = BOD, main = "BOD")
```



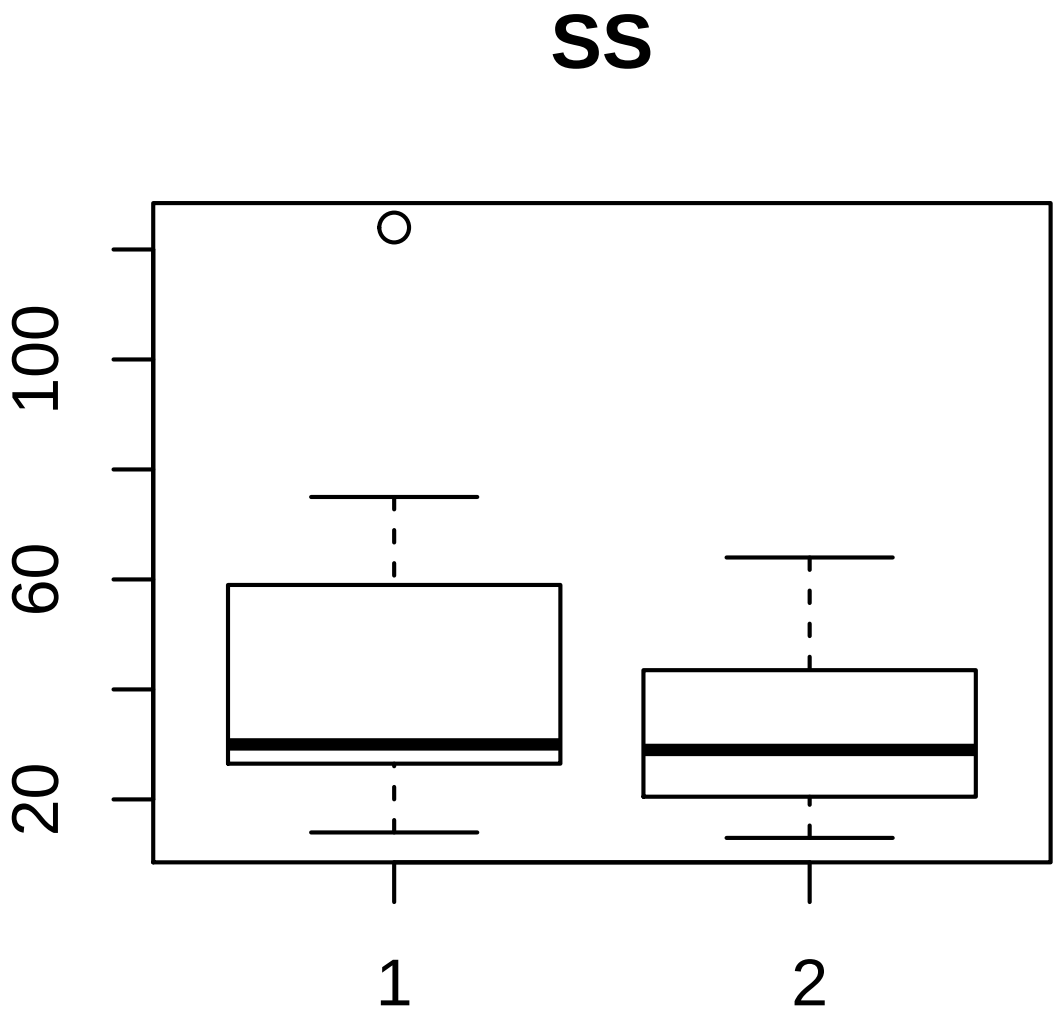
```r
head(SS)
```

```
#   Lab value
# 1   1    27
# 2   1    23
# 3   1    64
# 4   1    44
# 5   1    30
# 6   1    75
```

```r
tail(SS)
```

```
#    Lab value
# 17   2    64
# 18   2    30
# 19   2    64
# 20   2    56
# 21   2    20
# 22   2    21
```

```
boxplot(value ~ Lab, data = SS, main = "SS")
```



Univariate Paired Comparison

- Let X_{j1} denote the response to treatment 1 (or pre-test measurement), and let X_{j2} denote the response to treatment 2 (or post-test measurement).
- (X_{j1}, X_{j2}) are measurements recorded on the jth unit (pair). The n differences

$$D_j = X_{j1} - X_{j2}, \quad j = 1, 2, \ldots, n$$

reflect only the differential effects of the treatments.

Univariate Paired Comparison (cont)

- If $D_j \sim N(\delta, \sigma_d^2)$ where δ and σ_d^2 are the true mean and variances of the differences D_j, then the test statistic for testing whether $H_0 : \delta = 0$ vs $H_1 : \delta \neq 0$ is

$$t = \frac{\bar{D} - \delta_0}{s_d/\sqrt{n}} \sim t_{n-1} \quad (\text{reject } H_0 \text{ when} |t| > t_{n-1}(\alpha/2))$$

  where $\bar{D}$ and s_d^2 are the sample mean and sample variances of D_j.

Multivariate Paired Notations

We label the p responses within the jth unit as

$$
\begin{aligned}
X_{1j1} &= \text{variable 1 under treatment 1} \\
X_{1j2} &= \text{variable 2 under treatment 1} \\
&\vdots \quad \vdots \\
X_{1jp} &= \text{variable p under treatment 1} \\
X_{2j1} &= \text{variable 1 under treatment 2} \\
X_{2j2} &= \text{variable 2 under treatment 2} \\
&\vdots \quad \vdots \\
X_{2jp} &= \text{variable p under treatment 2}
\end{aligned}
$$

Multivariate Paired Differences

- The p paired-difference random variables become

$$\begin{aligned} D_{j1} &= X_{1j1} - X_{2j1} \\ D_{j2} &= X_{1j2} - X_{2j2} \\ &\vdots \quad \vdots \\ D_{jp} &= X_{1jp} - X_{2jp} \end{aligned}$$

- Let $\mathbf{D}_j' = [D_{j1}, D_{j2}, \ldots, D_{jp}]$, for $j = 1, 2, \ldots, n$, and $E(\mathbf{D}_j') = \delta' = [\delta_1, \delta_2, \ldots, \delta_p]$ and $Cov(\mathbf{D}_j) = \Sigma_d$.

Multivariate Paired Differences (cont)

- If $D_1, D_2, \ldots, D_n$ are IID $\sim N_p(\delta, \Sigma_d)$, inferences on δ can be based on the T^2-statistic

$$T^2 = n(\bar{\mathbf{D}} - \delta)'\mathbf{S}_d^{-1})(\bar{\mathbf{D}} - \delta) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

where $\bar{\mathbf{D}} = \frac{1}{n}\sum_{j=1}^{n} \mathbf{D}_j$ and

$$S_d = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{D}_j - \bar{\mathbf{D}})(\mathbf{D}_j - \bar{\mathbf{D}})'$$

Multivariate Paired Test and Confidence Region

- The condition $\delta = 0$ is equivalent to "no average difference between two treatments."
- $\bar{\mathbf{d}}$ and $\mathbf{S_d}$ are the realized values of $\bar{\mathbf{D}}$ and $\mathbf{S_D}$.
- An α-level test of $H_0 : \delta = 0$ vs $H_1 : \delta \neq 0$ for $N_p(\delta, \Sigma_d)$ population rejects H_0 if the observed

$$T^2 = n\bar{\mathbf{d}}\mathbf{S}_\mathbf{d}^{-1}\bar{\mathbf{d}} > \frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)$$

where $F_{p,n-p}(\alpha)$ is the upper (100α)th percentile of an F-distribution.

- A $100(1-\alpha)\%$ confidence region for δ consists of all δ such than

$$(\bar{\mathbf{d}} - \delta)' S_{\mathbf{d}}^{-1} (\bar{\mathbf{d}} - \delta) \leq \frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha)$$

Multivariate Paired Simultaneous Intervals

- A $100(1-\alpha)\%$ simultaneous confidence intervals for the individual mean differences δ_i are given by

$$\delta_i : \bar{d}_i \pm \sqrt{\frac{(n-1)}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_i}^2}{n}}$$

  where $\bar{d}_i$ is the ith element of $\bar{\mathbf{d}}$ and $s_{d_i}^2$ is the ith diagonal element of $\mathbf{S}_d$.

- The Bonferroni $100(1-\alpha)\%$ simultaneous confidence intervals for the individual mean differences are

$$\delta_i : \bar{d}_i \pm t_{n-1}\left(\frac{\alpha}{2p}\right) \sqrt{\frac{s_{d_i}^2}{n}}$$

Wastewater Treatment Example

```
# get paired differeces for BOD and SS
(d <- with(waste, cbind(BOD1-BOD2, SS1-SS2)))
```

```
#       [,1] [,2]
#  [1,]  -19   12
#  [2,]  -22   10
#  [3,]  -18   42
#  [4,]  -27   15
#  [5,]   -4   -1
#  [6,]  -10   11
#  [7,]  -14   -4
#  [8,]   17   60
#  [9,]    9   -2
# [10,]    4   10
# [11,]  -19   -7
```

Sample Mean and Covariances

```
n <- nrow(d)
p <- 2
(dbar <- colMeans(d))
```

```
# [1] -9.4 13.3
```

```
(S <- cov(d))
```

```
#      [,1] [,2]
# [1,]  199   88
# [2,]   88  419
```

Multivariate Paired Test

```
(T2 <- n*t(dbar)%*%solve(S)%*%dbar) # hotelling T2
```

```
#       [,1]
# [1,]   14
```

```
# scaled F critical value
p*(n-1)/(n-p)*qf(0.95,df1=p,df2=n-p)
```

```
# [1] 9.5
```

```r
# scale T2 so we can use original F-crit value
(T2.F <- T2/((n-1)*p/(n-p)))
```

```
#      [,1]
# [1,]  6.1
```

```r
# original F-critical value
(cval <- qf(0.95, df1 = p, df2 = n-p))
```

```
# [1] 4.3
```

```r
# We can also compute the p-value using
(pvalue <- 1 - pf((n-p)/(p*(n-1))*T2, df1=p,df2 = n-p))
```

```
#       [,1]
# [1,] 0.021
```

Multivariate Paired Test using ICSNP::HotellingsT2

```
ICSNP::HotellingsT2(d, mu = c(0, 0))
```

```
#
#   Hotelling's one sample T2-test
#
# data:  d
# T.2 = 6, df1 = 2, df2 = 9, p-value = 0.02
# alternative hypothesis: true location is not equal to c(0
```

Since $T2.F = 6.1 > 4.3, (p = 0.02)$, we reject H_0 at 5% level and conclude that there is a nonzero mean difference between the measurements of the two laboratories.

Simultaneous Confidence Intervals

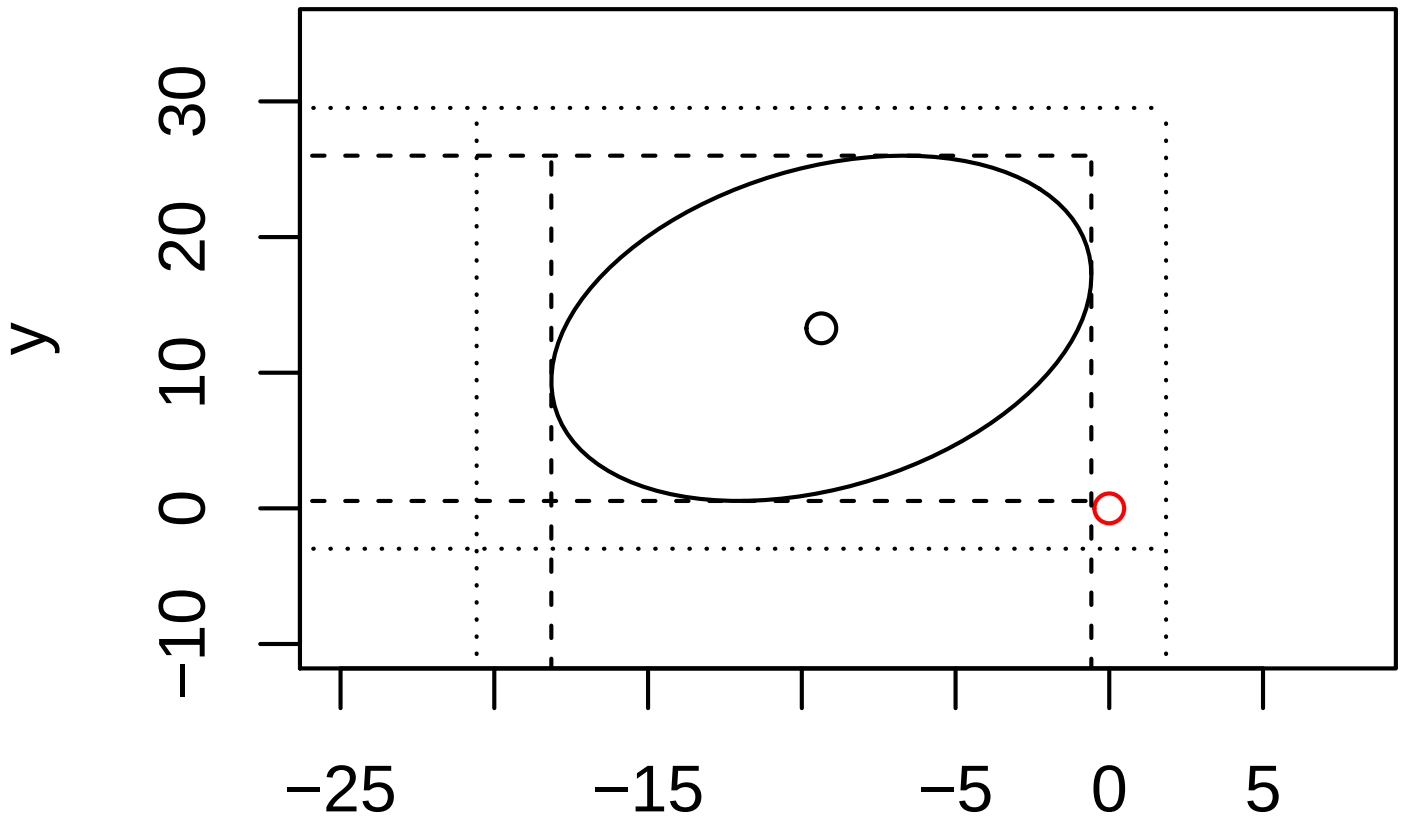
```
# The 95% simultaneous confidence intervals are
LL <- dbar - sqrt(cval)*sqrt(diag(S)/n)
UL <- dbar + sqrt(cval)*sqrt(diag(S)/n)

# Bonferroni Corrected simultaneous confidence t-intervals
ct <- qt(1-0.05/(2*2),df=n-1)
LLbc <- dbar - ct*sqrt(diag(S)/n)
ULbc <- dbar + ct*sqrt(diag(S)/n)
data.frame(LL_T2 = LL, UL_T2 = UL,
           LL_bc = LLbc, UL_bc = ULbc)
```

```
#    LL_T2 UL_T2 LL_bc UL_bc
# 1 -18.14 -0.58   -21   1.8
# 2   0.55 26.00    -3  29.5
```

Note that the 95% confidence intervals include zero, yet we reject H_0 (no difference) at 5% level. What is going on?

Plot the 95% confidence ellipse for the vector of mean differences



- Looking back at $T_2 = 14 > 9.47$, we have evidence that the confidence region excludes $\delta = 0$.
- The 95% simultaneous confidence coefficient $\sqrt{\frac{(n-1)}{(n-p)} F_{p,n-p}(\alpha)} = \sqrt{9.47} = 3.07$ applies to all possible linear combinations of the form $a_1\delta_1 + a_2\delta_2$.
- Our interval choices have $(a_1 = 1, a_2 = 0)$ and $(a_1 = 0, a_2 = 1)$ contains zero.

- However, other choices of a_1 and a_2 will produce simultaneous intervals that do not contain zero.
- If $H_0 : \delta = 0$ were not rejected, then all simultaneous interval would include zero.
- Note that in the boxplots, we see possible outliers in the data? Is the assumption of multivariate normality valid in this case?

```r
# outlier
waste[8,]
```

```
#   BOD1 SS1 BOD2 SS2
# 8   71 124   54  64
```

```r
# remove eight observation
ICSNP::HotellingsT2(d[-8, ], mu = c(0, 0))
```

```
#
#   Hotelling's one sample T2-test
#
# data:  d[-8, ]
# T.2 = 5, df1 = 2, df2 = 8, p-value = 0.04
# alternative hypothesis: true location is not equal to c(0
```

Repeated Measures Design (RMD) for Comparing Treatments

- q treatments (or time points) are compared with respect to a single response variable.
- Each subject or unit receives each treatment once over successive periods of time.
- The jth observation is $\mathbf{X}_j' = [X_{j1}, X_{j2}, \ldots, X_{jq}]$, $j = 1, 2, \ldots, n$.
- The repeated measures stem from the fact that all treatments are administered to each unit.

RMD for Comparing Treatments (cont)

- We consider contrasts of the components of $\mu = E(\mathbf{X}_j)$.

$$\begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \vdots \\ \mu_1 - \mu_q \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{bmatrix} = \mathbf{C}\mu$$

- The $q-1$ rows of C are linearly independent and each is a contrast vector.

Test of Equality of Treatments in RMD

Consider an $N_q(\mu, \Sigma)$ population, and let C be a contrast matrix. An α-level test of $H_0 : \mathbf{C}\mu = 0$ (equal treatment means) versus $H_1 : C\mu \neq 0$ is

$$T^2 = n(\mathbf{C}\bar{\mathbf{x}})'(\mathbf{CSC}')^{-1}(\mathbf{C}\bar{\mathbf{x}}) > \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,n-q+1}(\alpha)$$

where $F_{q-1,n-q+1}(\alpha)$ is the upper (100α)th percentile of an F-distribution with $q-1$ and $n-q+1$ d.f.;

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_j, \qquad \mathbf{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'$$

- A confidence region for contrast $C\mu$ is the set of all $C\mu$ such that

$$n(\mathbf{C}\bar{\mathbf{x}}-\mathbf{C}\mu)'(\mathbf{CSC}')^{-1}(\mathbf{C}\bar{\mathbf{x}}-\mathbf{C}\mu) \leq \frac{(n-1)(q-1)}{(n-q+1)}F_{q-1,n-q+1}(\alpha)$$

- Also, simultaneous $100(1-\alpha)\%$ confidence intervals for single contrasts $\mathbf{c}'\mu$ for any constrast vectors are

$$\mathbf{c}'\mu: \ \mathbf{c}'\bar{\mathbf{x}} \pm \sqrt{\frac{(n-1)(q-1)}{(n-q+1)}F_{q-1,n-q+1}(\alpha)}\sqrt{\frac{\mathbf{c}'\mathbf{Sc}}{n}}$$

Anesthetics Example

- Improved anesthetics are often developed by first studying their effects on animals.
- In one study, 19 dogs were initially given the drug pentobarbitol. Each dog was then administered CO_2 at each of two pressure levels.
- Next, halothane (H) was added, and the administration of CO_2 was repeated.

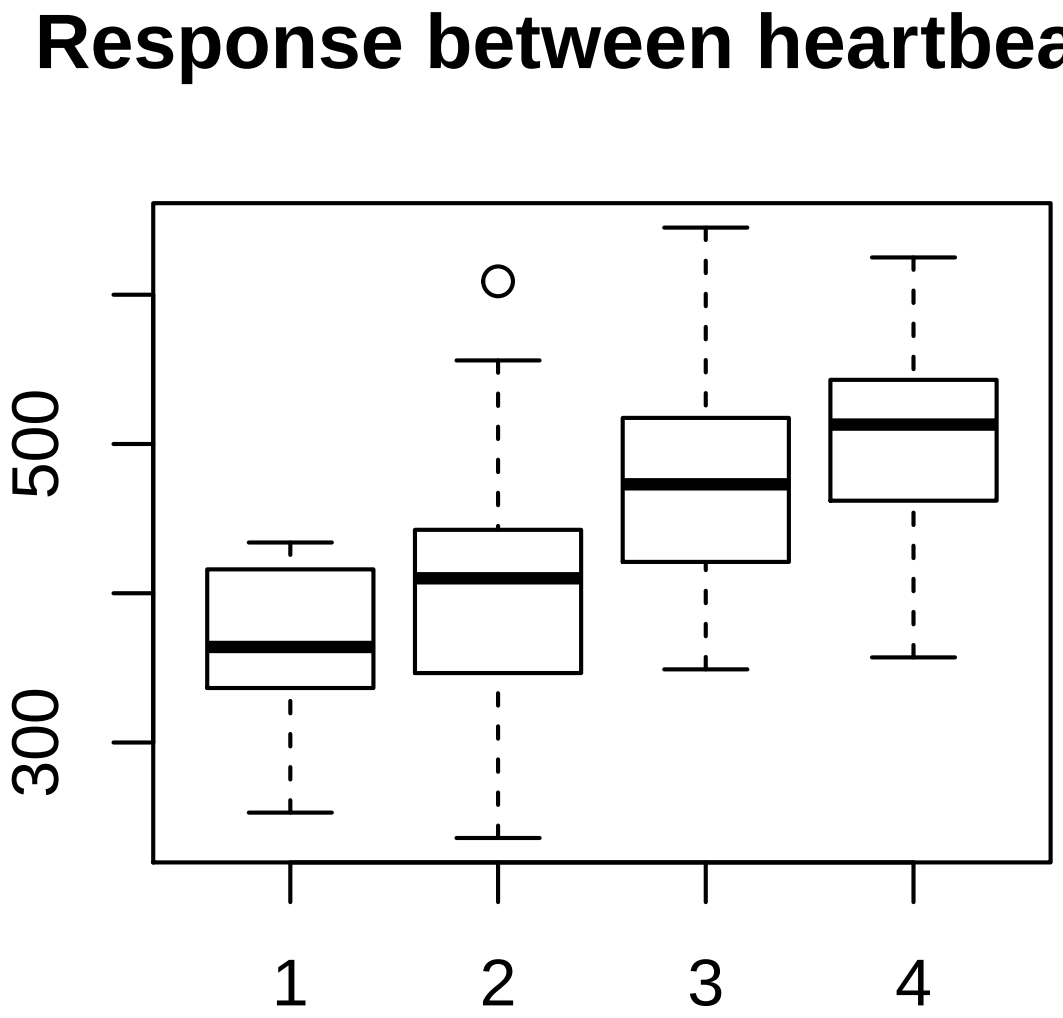
- The response, milliseconds between heartbeats, was measured for the four treatment combinations
  - Treatment 1 $=$ high CO_2 pressure without H
  - Treatment 2 $=$ low CO_2 pressure without H
  - Treatment 3 $=$ high CO_2 pressure with H
  - Treatment 4 $=$ low CO_2 pressure with H
- Analyze the anesthetizing effects of CO_2 pressure and halothene (H) from this repeated-measures design.

```r
time <- read.table("T6-2.dat")
colnames(time) <- c("T1","T2","T3","T4")
n <- nrow(time)
q <- ncol(time)
c(n,q)
```

```
# [1] 19  4
```

```r
heartbeats <- data.frame(treatment = c(rep(1,n),rep(2,n),
                                       rep(3,n),rep(4,n)),
                beats = with(time, c(T1, T2, T3, T4)))
```

```
boxplot(beats ~ treatment, data = heartbeats,
                main = "Response between heartbeats ")
```



Anesthetics Example : Treatment Contasts

- Let μ_1, μ_2, μ_3 and μ_4 denote the mean responses for treatments 1,2,3 and 4.
- There are three treatment contrasts that might be of interest in the experiment.
- $(\mu_3 + \mu_4) - (\mu_1 + \mu_2) =$ Halothane contrast representing the difference between the presence and absence of halothene
- $(\mu_1 + \mu_3) - (\mu_2 + \mu_4) = CO_2$ contrast representing the difference between high and low CO_2 pressure
- $(\mu_1 + \mu_4) - (\mu_2 + \mu_3) =$ Contrast representing the influence of halothane on CO_2 pressure differences ($H - CO_2$ pressure "interaction")

```r
# The contrast matrix is
c1 <- c(-1, -1, 1, 1)
c2 <- c(1, -1, 1, -1)
c3 <- c(1, -1, -1, 1)
(C <- rbind(c1,c2,c3))
```

```
#     [,1] [,2] [,3] [,4]
# c1   -1   -1    1    1
# c2    1   -1    1   -1
# c3    1   -1   -1    1
```

```r
(xbar <- colMeans(time))
```

```
#  T1  T2  T3  T4
# 368 405 479 503
```

```
(S <- cov(time))
```

```
#       T1   T2   T3   T4
# T1 2819 3568 2943 2295
# T2 3568 7963 5304 4065
# T3 2943 5304 6851 4500
# T4 2295 4065 4500 4879
```

```
(T2 <- n*t(C%*%xbar)%*%solve(C%*%S%*%t(C))%*%C%*%xbar)
```

```
#       [,1]
# [1,]  116
```

```
(cval <- (n-1)*(q-1)/(n-q+1) * qf(0.95,
                                  df1=q-1,df2 = n-q+1))
```

```
# [1] 11
```

```r
# scaled T2, original F crit
c(T2/((n-1)*(q-1)/(n-q+1)),
  qf(0.95, df1=q-1,df2 = n-q+1))
```

```
# [1] 34.4  3.2
```

```r
# We can also compute the p-value using
(pvalue <- 1 - pf((n-q+1)/((n-1)*(q-1))*T2,
    df1=q-1,df2 = n-q+1))
```

```
#           [,1]
# [1,] 3.3e-07
```

Tests of Equality of Treatments in RMD using ICSNP::HotellingsT2

```r
# transform data using contrasts matrix
time.C <- t(C %*% t(time))
ICSNP::HotellingsT2(time.C, mu = c(0, 0, 0))
```

```
#
#   Hotelling's one sample T2-test
#
# data:  time.C
# T.2 = 30, df1 = 3, df2 = 20, p-value = 3e-07
# alternative hypothesis: true location is not equal to c(0
```

Strong evidence in the sample that the vector of constrasts is different from the zero vector. This means that some, if not all, of the contrasts values are different from zero.

- To see which of the contrasts are responsible for the rejection of H0.
- Let's perform two-sided t-test on each contrasts. We can also get equivalent result if we instead construct confidence intervals for the contrasts.
- We use the base function `t.test()` to do this.
- Since there three contrasts, we use Bonferonni correction on the level of significance.
- Bonferroni corrected level of significance: $\alpha/3$

```
(alpha0 <- 0.05/3)  # Bonferonni correction
```

```
# [1] 0.017
```

Contrast 1: $(\mu_3 + \mu_4) - (\mu_1 + \mu_2)$

Halothane contrast representing the difference between the presence and absence of halothene

```
time.c1 <- t(c1 %*% t(time))
t.test(time.c1, mu = 0, conf.level = 1 - alpha0)
```

```
#
#   One Sample t-test
#
# data:  time.c1
# t = 9, df = 20, p-value = 2e-08
# alternative hypothesis: true mean is not equal to 0
# 98 percent confidence interval:
#  151 268
# sample estimates:
# mean of x
#       209
```

- It strongly seems that there is a halothene effect.
- The presence of halothene produces longer times between heartbeats.

Contrast 2: $(\mu_1 + \mu_3) - (\mu_2 + \mu_4)$

CO_2 contrast representing the difference between high and low CO_2 pressure

```
time.c2 <- t(c2 %*% t(time))
t.test(time.c2, mu = 0, conf.level = 1 - alpha0)
```

```
#
#   One Sample t-test
#
# data:  time.c2
# t = -4, df = 20, p-value = 0.002
# alternative hypothesis: true mean is not equal to 0
# 98 percent confidence interval:
#  -104  -16
# sample estimates:
# mean of x
#       -60
```

- The result of the test on the second contrast indicates that there is an effect due to CO_2 pressure.
- Lower CO_2 pressure produces longer times between heartbeats.

Contrast 3: $(\mu_1 + \mu_4) - (\mu_2 + \mu_3)$

Representing the influence of halothane on CO_2 pressure differences ($H - CO_2$ pressure "interaction")

```
time.c3 <- t(c3 %*% t(time))
t.test(time.c3, mu = 0, conf.level = 1 - alpha0)
```

```
#
#   One Sample t-test
#
# data:  time.c3
# t = -0.6, df = 20, p-value = 0.5
# alternative hypothesis: true mean is not equal to 0
# 98 percent confidence interval:
#  -65  40
# sample estimates:
# mean of x
#       -13
```

- No significant interaction effect between Halothen and CO_2 is detected.
- Halothene produces longer times between heartbeats at both levels of CO_2 pressure.
- Some caution must be exercised in our interpretation because the trials with Halothene must follow those without (meaning there is a possible time trend).